

## 1

(a)

i: Sannolikheten att ett visst element  $i$  kommer med i stickprovet.

ii:

$$\pi_{i,OSU \text{ u.ä.}} = \frac{n}{N}, \text{ där } N - \text{populationsstorlek, } n - \text{stickprovsstorlek}$$

iii:

$$\pi_{i,SY} = \frac{1}{k} = \frac{n}{N}$$

iv:

$$\pi_{i,STOSU \text{ u.ä.}} = \frac{n_l}{N_l}, \text{ där } N_l - \text{populationsstorlek i stratum } l, \\ n_l - \text{stickprovsstorlek allokerat till stratum } l$$

v:

$$\pi_{i,STSY} = \frac{1}{k_l} = \frac{n_l}{N_l}$$

(b)

$$\hat{V}(\hat{\mu}) = \frac{N-n}{N} \frac{s^2}{n} \stackrel{\substack{\uparrow \\ \text{Från formelblad et,} \\ \text{OSU u.ä. med HT-est.}}}{=} \frac{N-n}{N} \frac{\sum_{i=1}^n y_i^2 - n\bar{y}^2}{n(n-1)} \stackrel{\substack{\uparrow \\ \text{Bernoullivariabel}}}{=} \frac{N-n}{N} \frac{\sum_{i=1}^n y_i - n\bar{y}^2}{n(n-1)} =$$

$$= \frac{N-n}{N} \frac{n\bar{y} - n\bar{y}^2}{n(n-1)} = \frac{N-n}{N} \frac{n\bar{y}(1-\bar{y})}{n(n-1)} = \frac{N-n}{N} \frac{n\hat{p}(1-\hat{p})}{n(n-1)} = \frac{N-n}{N} \frac{\hat{p}\hat{q}}{n-1}$$

## 2

(a)

En erbjuden skala som innehåller svarsalternativ kan vara starkt ledande i vissa sammanhang (se t.ex. OH från F2, s. 22), varför ur den synvinkeln är alt. A att föredra. Öppna svar är däremot svårare att koda eller kan leda till att respondentens missförståelse av frågan uppenbarar sig (vilket inte syns vid alt. B). Det gäller att avväga mellan dessa påföljder, men vill man undvika mätfel tallar allt om att alternativet A ska väljas.

(b)

Att komma ihåg saker exakt kräver ansträngning, varför vi vill—även i sammanhang av en pappersblankett—lata respondenten förstå att hon bör ta tid på sig att svara på denna fråga; alltså, vi ska ta ett långsammare tempo. Dessutom är det av vikt att respondenten ska ta hänsyn till vår referensperiod (dvs vi måste minimera telescoping-effekten, vilket är det fenomen då respondenten flyttar referensperioden bakåt eller framåt i tiden); därför ska vi betona referensperioden samt även stödja minnet genom t ex att ange några exakta datum. En lämplig början på frågan kan alltså vara (med senaste terminen som referensperioden):

*”Tänk dig förra terminen, alltså den termin som började 21 augusti 2002 och som slutade 15 januari 2003. På tentor och omtentor för kurser som du läste under den terminen, hur många VG sammanlagt har du fått?”*

## 3

Pop:  $N = 968$  ensamboende personer.

Usv: antal rum i bostaden,  $y$ .

Param:

(a) total,

$$\tau_y = \sum_{i=1}^N y_i = N \frac{\sum_{i=1}^N y_i}{N} = N\mu_y$$

(b) total i undergrupp,

$$\tau_{My} = \sum_{i=1}^{N_M} y_{Mi} = N_M \frac{\sum_{i=1}^{N_M} y_{Mi}}{N_M} = N_M \mu_M$$

Urv: OSU u.å.,  $n = 20$

Paramsk:

(a)

$$\hat{\tau}_y \underset{\substack{= \\ \uparrow \\ HT\text{-}estimatorn}}{=} \sum_{i=1}^n \frac{y_i}{\pi_i} = \sum_{i=1}^n \frac{y_i}{\frac{n}{N}} = N \frac{\sum_{i=1}^n y_i}{n} = N\bar{y}.$$

(b)

$$\hat{\tau}_{My} = N_M \bar{y}_M = N_M \frac{\sum_{i=1}^{n_M} y_{Mi}}{n_M} \underset{\substack{= \\ \uparrow \\ N_M \text{ ej känt,} \\ \frac{N_M}{n_M} \text{ ersätts med } \frac{N}{n}}}{=} N \frac{\sum_{i=1}^{n_M} y_{Mi}}{n}.$$

Varsk:

(a)

$$\hat{V}(\hat{\tau}_y) = \hat{V}(N\bar{y}) = N^2 \hat{V}(\bar{y}) \underset{\substack{= \\ \uparrow \\ \text{Från formelblad et,} \\ \text{OSU u.å. med HT-est.}}}{=} N^2 \frac{N-n}{N} \frac{s_y^2}{n}$$

(b)

$$\hat{V}(\hat{\tau}_{My}) \underset{\substack{= \\ \uparrow \\ \text{Från formelblad et,} \\ \text{undergrupp, fallet } N_M \text{ ej känt}}}{=} N^2 \frac{N-n}{N} \frac{\sum_{j=1}^{n_M} y_{Mj}^2 - \frac{(\sum_{j=1}^{n_M} y_{Mj})^2}{n}}{n(n-1)}$$

## Data

Obs	Kön	Inkomst	Ant. rum	Rum(M)	(Rum(M))^2
1	M	18	2	2	4
2	M	16	1	1	1
3	M	23	3	3	9
4	M	17	1	1	1
5	M	20	2	2	4
6	M	28	4	4	16
7	M	19	2	2	4
8	M	32	3	3	9
9	M	22	1	1	1
10	K	17	1	0	0
11	K	16	1	0	0
12	K	19	2	0	0
13	K	24	2	0	0
14	K	21	3	0	0
15	K	19	1	0	0
16	K	18	1	0	0
17	K	22	3	0	0
18	K	17	1	0	0
19	K	20	1	0	0
20	K	19	2	0	0
Summa		407	37	19	49
s^2=			0.8711	1.6289	

(a)

$$\hat{\tau}_y = N \frac{\sum_{i=1}^n y_i}{n} = 968 \times \frac{37}{20} \approx 1791$$

$$\hat{V}(\hat{\tau}_y) = N^2 \frac{N-n}{N} \frac{s_y^2}{n} = 968^2 \times \frac{968-20}{968} \times \frac{0.8711}{20} = 9.3702 \times 10^5 \times 0.97934 \times 4.3555 \times 10^{-2} = 39969$$

$$\text{Standardavvikelse, } \sqrt{\hat{V}(\hat{\tau}_y)} = \sqrt{39969} \approx 200$$

(b)

$$\hat{\tau}_{My} = N \frac{\sum_{i=1}^n y_{Mi}}{n} = 968 \times \frac{19}{20} \approx 920$$

$$\hat{V}(\hat{\tau}_{My}) = N^2 \frac{N-n}{N} \frac{\sum_{j=1}^n y_{Mj}^2 - \frac{(\sum_{j=1}^n y_{Mj})^2}{n}}{n(n-1)} = 968^2 \times \frac{968-20}{968} \times \frac{49 - \frac{19^2}{20}}{20(20-1)} =$$

$$= 9.3702 \times 10^5 \times 0.97934 \times \frac{1}{20} \times 1.6289 = 74739$$

$$\text{Standardavvikelse, } \sqrt{\hat{V}(\hat{\tau}_{My})} = \sqrt{74739} \approx 273.4$$

4

Pop:  $N = 968$  ensamboende personer, varav  $N_M = 440$  män och  $N_K = 528$  kvinnor; alltså, antal efterstrata  $L = 2$ .

Usv: antal rum i bostaden,  $y$ .

Param: total,

$$\tau_y = \sum_{l=1}^L \tau_l = \sum_{l=1}^L \sum_{i=1}^{N_l} y_{li} = \sum_{l=1}^L N_l \frac{\sum_{i=1}^{N_l} y_{li}}{N_l} = \sum_{l=1}^L N_l \mu_l$$

Urv: OSU u.å.,  $n = 20$ ,  $n_M = 9$ ,  $n_K = 11$ .

Paramsk:

$$\hat{\tau}_{y,post} \underset{\substack{\uparrow \\ \text{efterstratifiering}}}{=} \sum_{l=1}^L N_l \bar{y}_l$$

Varsk:

$$\hat{V}(\hat{\tau}_{y,post}) = \hat{V}(N\bar{y}_{post}) = N^2 \hat{V}(\bar{y}_{post}) \underset{\substack{\uparrow \\ \text{Approx. med } \hat{V}(\bar{y}_{st}), \\ \text{från formelblad et}}}{\approx} N^2 \frac{1}{N^2} \sum_{l=1}^L N_l^2 \frac{N_l - n_l}{N_l} \frac{s_{ly}^2}{n_l} = \sum_{l=1}^L N_l^2 \frac{N_l - n_l}{N_l} \frac{s_{ly}^2}{n_l}$$

Data

Obs	Kön	Inkomst	Ant. rum	Rum(M)	Rum(K)
1	M	18	2	2	
2	M	16	1	1	
3	M	23	3	3	
4	M	17	1	1	
5	M	20	2	2	
6	M	28	4	4	
7	M	19	2	2	
8	M	32	3	3	
9	M	22	1	1	
10	K	17	1		1
11	K	16	1		1
12	K	19	2		2
13	K	24	2		2
14	K	21	3		3
15	K	19	1		1
16	K	18	1		1
17	K	22	3		3
18	K	17	1		1
19	K	20	1		1
20	K	19	2		2
Summa		407	37	19	18
s^2=				1.1111	0.6545

$$\hat{\tau}_{y,post} = \sum_{l=1}^L N_l \bar{y}_l = N_M \bar{y}_M + N_K \bar{y}_K = 440 \times \frac{19}{9} + 528 \times \frac{18}{11} = 928.89 + 864.0 \approx 1793$$

(Kommentar: andelen män i stickprovet,  $\frac{n_M}{n} = \frac{9}{20} = 0.45$ , skilde sig lite från andelen män i populationen,  $\frac{N_M}{N} = \frac{440}{968} = 0.4545$ , varför den ovägda totalskattningen i uppg. 3a påverkades lite av efterstratifieringen.)

$$\hat{V}(\hat{\tau}_{y,post}) = \sum_{l=1}^L N_l^2 \frac{N_l - n_l}{N_l} \frac{s_{ly}^2}{n_l} = N_M^2 \frac{N_M - n_M}{N_M} \frac{s_{My}^2}{n_M} + N_K^2 \frac{N_K - n_K}{N_K} \frac{s_{Ky}^2}{n_K} =$$

$$= 440^2 \times \frac{440-9}{440} \times \frac{1.1111}{9} + 528^2 \times \frac{528-11}{528} \times \frac{0.6545}{11} = 1.936 \times 10^5 \times 0.97955 \times 0.12346 +$$

$$+ 2.7878 \times 10^5 \times 0.97917 \times 0.0595 = 23413 + 16242 = 39655$$

$$\text{Standardavvikelse, } \sqrt{\hat{V}(\hat{\tau}_{y,post})} = \sqrt{39655} \approx 199$$

## 5

Pop:  $N = 968$  ensamboende personer, varav  $N_M = 440$  män och  $N_K = 528$  kvinnor; alltså, antal efterstrata  $L = 2$ .

Usv: antal rum i bostaden,  $y$ .

Hjv: inkomst,  $x$ .  $\mu_{x,M} = 22$ ,  $\mu_{x,K} = 18$ .

Param: total,

$$\tau_y = \sum_{i=1}^N y_i$$

eller

$$\tau_y = \sum_{l=1}^L \tau_l = \sum_{l=1}^L \sum_{i=1}^{N_l} y_{li} = \sum_{l=1}^L N_l \frac{\sum_{i=1}^{N_l} y_{li}}{N_l} = \sum_{l=1}^L N_l \mu_l$$

Urv: OSU u.å.,  $n = 20$ ,  $n_M = 9$ ,  $n_K = 11$ .

Kan lösas, med utnyttjande av hjälpinformation, på två sätt: (a) en regressionslinje för hela populationen skattas, vilket är det tänkta sättet att lösa uppgiften (b) en regressionslinje per efterstratum i populationen skattas, vilket också är rätt men mer komplext att beräkna.

(a)

Paramsk:

$$\hat{\tau}_{y,r} \underset{\substack{= \\ \uparrow \\ \text{kvotskattning}}}{=} N r \mu_x = N \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i} \mu_x, \text{ där } \mu_x = \frac{N_M \mu_{x,M} + N_K \mu_{x,K}}{N}$$

Varsk:

$$\hat{V}(\hat{\tau}_{y,r}) = \hat{V}(N \hat{\mu}_{y,r}) = N^2 \hat{V}(\hat{\mu}_{y,r}) \underset{\substack{= \\ \uparrow \\ \text{Från formelbladet,} \\ \text{kvotskattning}}}{=} N^2 \frac{N-n}{nN} \frac{\sum_{i=1}^n (y_i - r x_i)^2}{n-1} = N^2 \frac{N-n}{N} \frac{s_r^2}{n}$$

(b)

Paramsk:

$$\hat{\tau}_{y,r,\text{post}} = \sum_{l=1}^L \hat{\tau}_{y,l,r} = \sum_{l=1}^L N_l \hat{\mu}_{y,l,r} \underset{\substack{= \\ \uparrow \\ \text{kvotskattning}}}{=} \sum_{l=1}^L N_l r_l \mu_{x,l} = \sum_{l=1}^L N_l \frac{\sum_{i=1}^{n_l} y_{i,l}}{\sum_{i=1}^{n_l} x_{i,l}} \mu_{x,l}$$

Varsk:

$$\begin{aligned}\hat{V}(\hat{\tau}_{y,r,post}) &= \sum_{l=1}^L \hat{V}(\hat{\tau}_{y,l,r}) = \sum_{l=1}^L \hat{V}(N_l \hat{\mu}_{y,l,r}) = \sum_{l=1}^L N_l^2 \hat{V}(\hat{\mu}_{y,l,r}) \\ &\stackrel{\substack{\uparrow \\ \text{Från formelbladet,} \\ \text{kvotskattning}}}{=} \\ &= \sum_{l=1}^L N_l^2 \frac{N_l - n_l}{n_l N_l} \frac{\sum_{i=1}^{n_l} (y_{i,l} - r_l x_{i,l})^2}{n_l - 1} = \sum_{l=1}^L N_l^2 \frac{N_l - n_l}{N_l} \frac{s_{r,l}^2}{n_l}\end{aligned}$$

Data

Obs	Kön	Inkomst	Ink(M)	Ink(K)	Ant. rum	Ant(M)	Ant(K)	(y-rx)^2	(y-rx(M))^2	(y-rx(K))^2
1	M	18	18		2	2		0.1322	0.0606	
2	M	16	16		1	1		0.2066	0.3125	
3	M	23	23		3	3		0.8264	0.5760	
4	M	17	17		1	1		0.2975	0.4309	
5	M	20	20		2	2		0.0331	0.0026	
6	M	28	28		4	4		2.1157	1.6175	
7	M	19	19		2	2		0.0744	0.0221	
8	M	32	32		3	3		0.0083	0.0139	
9	M	22	22		1	1		1.0000	1.3078	
10	K	17		17	1		1	0.2975		0.1966
11	K	16		16	1		1	0.2066		0.1285
12	K	19		19	2		2	0.0744		0.1496
13	K	24		24	2		2	0.0331		0.0014
14	K	21		21	3		3	1.1901		1.4810
15	K	19		19	1		1	0.5289		0.3760
16	K	18		18	1		1	0.4050		0.2791
17	K	22		22	3		3	1.0000		1.2816
18	K	17		17	1		1	0.2975		0.1966
19	K	20		20	1		1	0.6694		0.4874
20	K	19		19	2		2	0.0744		0.1496
Summa		407	195	212	37	19	18	9.4711	4.3439	4.7275
r=		0.090909	0.097436	0.084906						
s^2=								0.4985	0.5430	0.4727

(a)

$$\hat{\tau}_{y,r} = Nr\mu_x = N \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i} \mu_x = 968 \times 0.09091 \times \frac{440 \times 22 + 528 \times 18}{968} = 968 \times 0.09091 \times 19.818 = 1744.0$$

$$\hat{V}(\hat{\tau}_{y,r}) = N^2 \frac{N-n}{nN} \frac{\sum_{i=1}^n (y_i - rx_i)^2}{n-1} = N^2 \frac{N-n}{N} \frac{s_r^2}{n} = 968^2 \times \frac{968-20}{968} \times \frac{0.4985}{20} = 968^2 \times 0.97934 \times 2.4925 \times 10^{-2} = 22873$$

$$\text{Standardavvikelse, } \sqrt{\hat{V}(\hat{\tau}_{y,r})} = \sqrt{22873} = 151.24$$

(b)

$$\begin{aligned}\hat{\tau}_{y,r,post} &= \sum_{l=1}^L N_l \frac{\sum_{i=1}^{n_l} y_{i,l}}{\sum_{i=1}^{n_l} x_{i,l}} \mu_{x,l} = N_M \frac{\sum_{i=1}^{n_M} y_{i,M}}{\sum_{i=1}^{n_M} x_{i,M}} \mu_{x,M} + N_K \frac{\sum_{i=1}^{n_K} y_{i,K}}{\sum_{i=1}^{n_K} x_{i,K}} \mu_{x,K} = \\ &= 440 \times \frac{19}{195} \times 22 + 528 \times \frac{18}{212} \times 18 = 943.18 + 806.94 \approx 1750\end{aligned}$$

$$\begin{aligned}\hat{V}(\hat{\tau}_{y,r,post}) &= \sum_{l=1}^L N_l^2 \frac{N_l - n_l}{N_l} \frac{s_{r,l}^2}{n_l} = N_M^2 \frac{N_M - n_M}{N_M} \frac{s_{r,M}^2}{n_M} + N_K^2 \frac{N_K - n_K}{N_K} \frac{s_{r,K}^2}{n_K} = \\ &= 440^2 \times \frac{440-9}{440} \times \frac{0.5430}{9} + 528^2 \times \frac{528-11}{528} \times \frac{0.4727}{11} = 11442 + 11731 = 23173\end{aligned}$$

$$\text{Standardavvikelse, } \sqrt{\hat{V}(\hat{\tau}_{y,r,post})} = \sqrt{23173} = 152.23$$